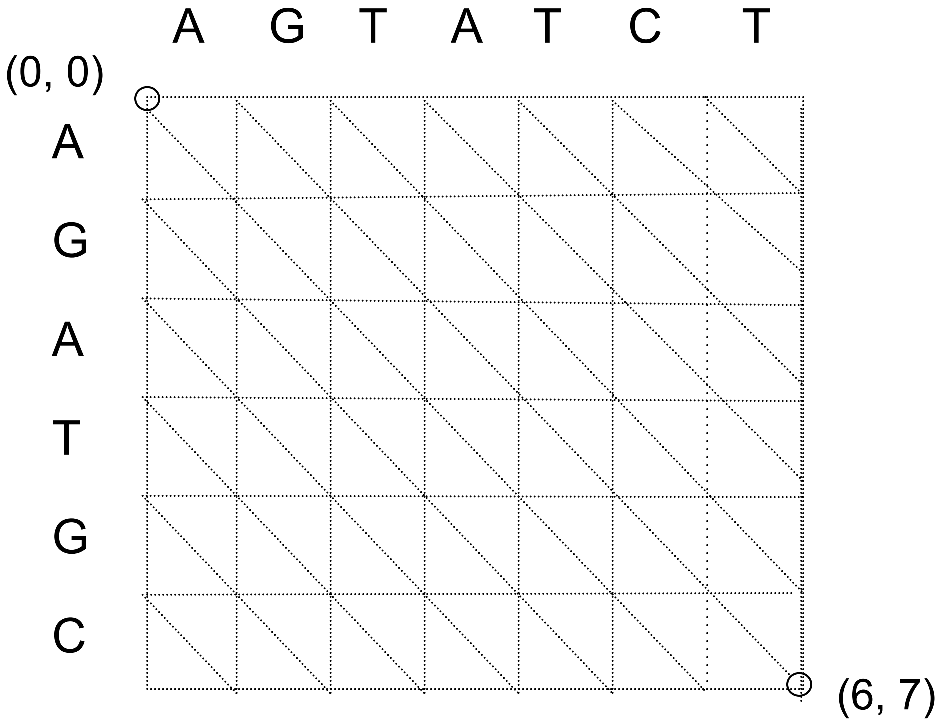
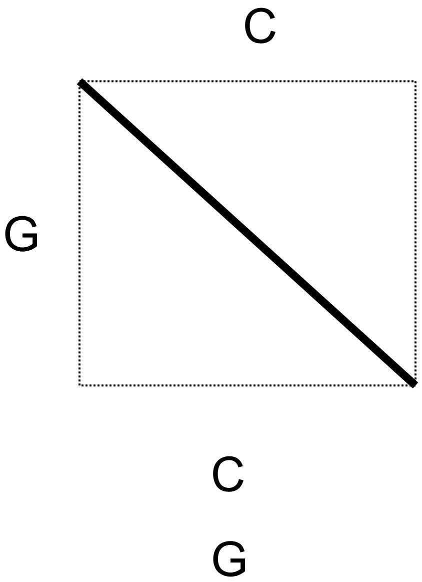
4. Sequence comparison

- How to predict gene/protein functions?
- Why gene/protein sequences may be similar?
- Measuring sequence similarity
- Aligning sequences is an optimization problem
- **A sequence alignment as a path**
- A path is optimal if all its sub-paths are optimal
- Alignment costs
- A recursive algorithm
- Recursion can be avoided: an iterative version
- How efficient is this algorithm?

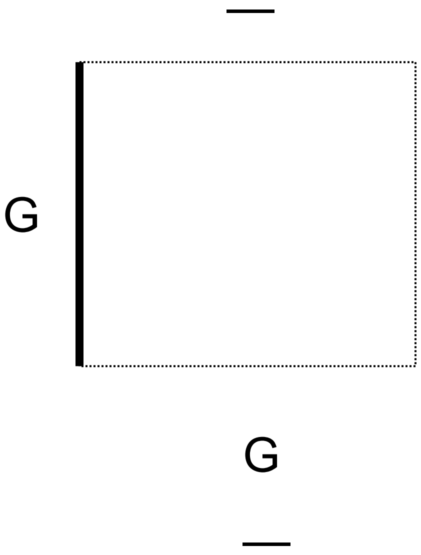
Second sequence

First sequence

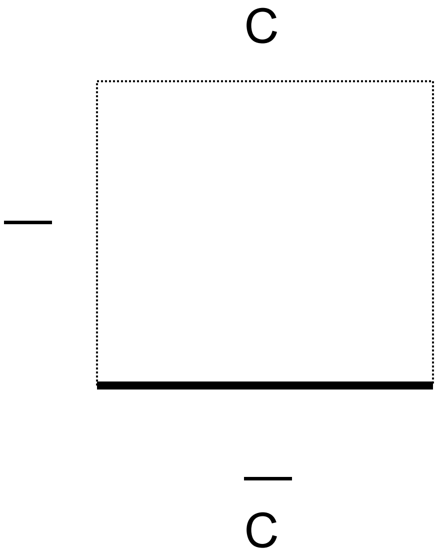




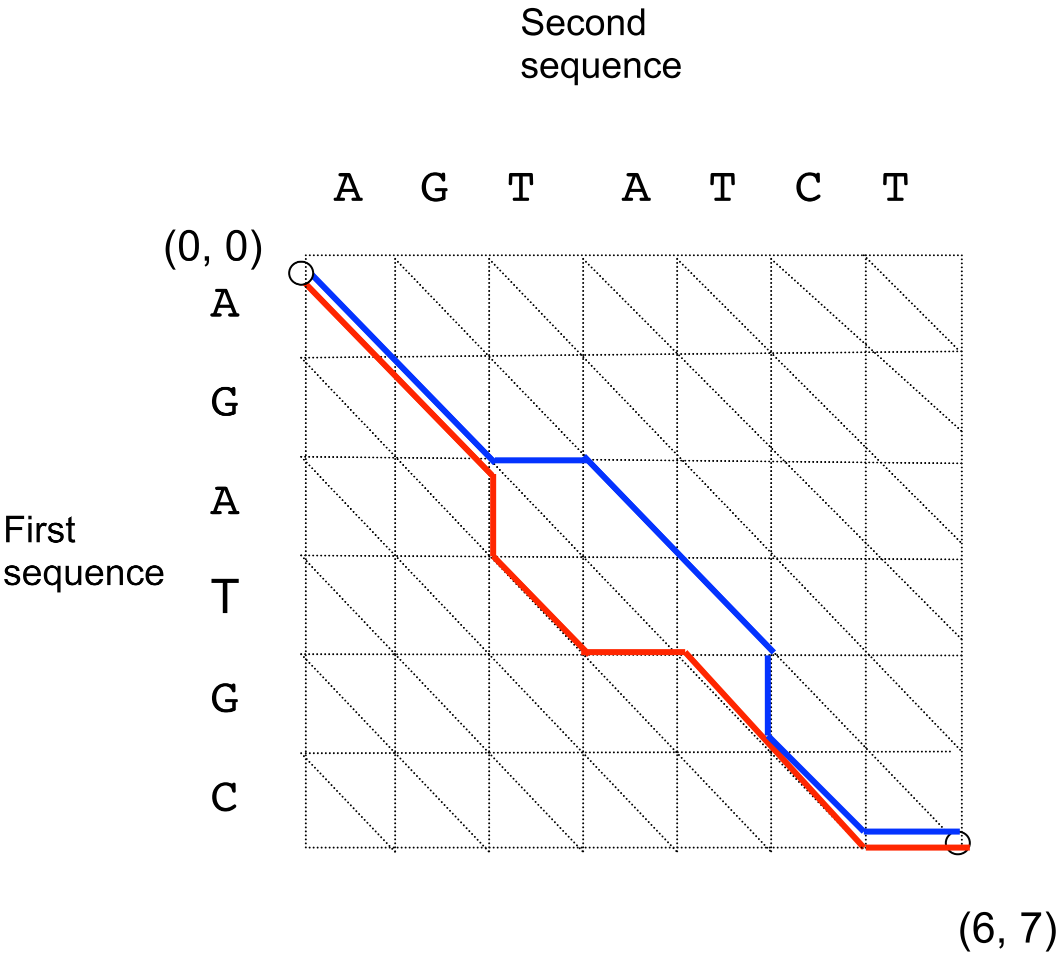
Substitution



Insertion in the second sequence (or deletion in the first)



Insertion in the first sequence (or deletion in the second)



A G — A T G C —
A G T A T — C T

A G A T — G C —
A G — T A T C T